



Al-Samman, I., Doufexi, A., & Beach, M. (2016). A C-RAN Architecture for LTE Control Signalling. In *2016 IEEE 83rd Vehicular Technology Conference (VTC Spring): Proceedings of a meeting held 15-18 May 2016, Nanjing, China* [7504061] Institute of Electrical and Electronics Engineers (IEEE).
<https://doi.org/10.1109/VTCSpring.2016.7504061>

Peer reviewed version

Link to published version (if available):
[10.1109/VTCSpring.2016.7504061](https://doi.org/10.1109/VTCSpring.2016.7504061)

[Link to publication record in Explore Bristol Research](#)
PDF-document

This is the author accepted manuscript (AAM). The final published version (version of record) is available online via IEEE at <http://ieeexplore.ieee.org/document/7504061/>. Please refer to any applicable terms of use of the publisher.

University of Bristol - Explore Bristol Research

General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:
<http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/>

A C-RAN Architecture for LTE Control Signalling

Imad Al-Samman, Angela Doufexi and Mark Beach

Communication Systems & Networks Group, Department of Electrical and Electronic Engineering
University of Bristol, Bristol, United Kingdom

Email: {[Imad.Al-samman](mailto:Imad.Al-samman@bristol.ac.uk), [M.A.Beach](mailto:M.A.Beach@bristol.ac.uk), [A.Doufexi](mailto:A.Doufexi@bristol.ac.uk)}@bristol.ac.uk

Abstract- The current LTE network architecture experiences high level of signalling traffic between control plane entities. As a result many innovative architectures have been proposed including SDN to simplify the network. Cloud RAN on the other hand proposes a pioneering paradigm in terms of sustaining the profitability margin with unprecedented surge of mobile traffic and providing better performance to the end users. Schemes have been proposed to compute and analyse the signalling load in the new SDN LTE architecture; however in our best understanding, none of them consider future CRAN architectures. Thus in this paper we propose a new SDN CRAN architecture and present an analysis of the signalling load, evaluate the performance and compare it against existing architectures in the literature. Evaluation results show significant improvement of the proposed schemes in terms of reduction in the signalling load when addressing several network metrics.

Index Terms—C-RAN; SDN; LTE, Control signalling.

I. INTRODUCTION

The immense surge of smart devices, new services and applications results in high traffic loads for the current networks. However the current cellular network architecture cannot cope with the unprecedented rate of growth in network use. Furthermore, a related study alerts the ending of the era of mobile network's profitability in 2015 [1]. Therefore the need to reshape the networks architecture and capabilities becomes vital for both reducing cost and enhancing the new services revenue. The broad range of mobile applications in smart phones and their corresponding keep-alive signalling would cause a major challenge for networks in respect of increasing the load of LTE signalling to keep up with the short messages generated by those applications. Operators need to come up with new approaches to face the aforementioned difficulties. The emerge of Software Define Network (SDN) [2] as a networking archetype in wired networks through its OpenFlow (OF) protocol and attractive features in separating the data and control planes has inspired researchers in both academia and industry to develop and deploy this architecture for wireless networks. In this context there have been many recent studies on the advantages of SDN in LTE [3][4][5]. However the previous studies have only considered the D-RAN (Distributed Radio Access Network) architecture. This paper will investigate the potential performance gain of deploying the cloud RAN (C-RAN) and compare it against D-RAN related studies. The authors in [6] have acknowledged the significance of using C-RAN in a handover context, where the signalling overhead is expected to increase due to the deployment of multi-tier cellular networks. Therefore, taking C-RAN into account along with the SDN approaches will help in terms of

reducing the overall signalling and simplifying the network topology from the controller perspective.

In this paper we present new architectures based on what is stated above and perform the overall signalling load analysis. The objective is to determine a better architecture in terms of lower signalling load. The related analysis will address multiple network parameters such as cell area and tracking area update. The rest of the paper is organized as follow: section II describes the calculation of the maximum distance between the Radio Remote Head (RRH) and Base Band Unit (BBU). The newly proposed network architecture with the corresponding mathematical modelling of different network metrics is introduced in section III. Section IV evaluates the performance of the new schemes and compares them with previous schemes. Finally we conclude the paper in section V.

II. C-RAN SIZE "PROBLEM FORMULATION"

Despite the benefits of using C-RAN, the main challenge is the maximum radius of a C-RAN that a BBU can manage. The UE in a standard LTE network should receive ACK/NACK from eNB in three sub-frames after sending uplink data to comply with the HARQ protocol. Hence the eNB needs to finish the DL processing within 3 ms after receiving the UL data. In C-RAN this timing requirement is difficult to meet due to the transmission line of fibre optics. This new medium links the RRH and the BBU which functions as the eNB. In addition, there is processing delay introduced by the active equipment in the front-haul links. According to [7], in order to meet the timing constraint in the standards the vendors need to amend the BBU design to accelerate the DL process and shorten it to be 2.7 ms. The aim of delay reduction is to compensate the latency that occurs due to the newly added separation distance between the RRH and the BBU. Overall transmission delay in the UL and its ACK/NACK response is presented in [7]. Table I illustrates the required delay values at all network components mentioned previously to satisfy the overall delay requirements of 3 ms. The Max fibre Round Trip Time (RTT) is given by: $3\text{ ms} - (A + B + C + D) = 246\text{ }\mu\text{s}$ where A, B, C and D are defined in Table I. By taking into account that the transmission latency on fibre is 5µsec/km, the Max fibre distance = $\frac{\text{Max Fibre RTT}}{2 \times 5\text{ }\mu\text{sec/km}} = 24.6\text{ km}$. We assume the standard cell radius r_c in LTE is 4 km through the rest of the study and a hexagonal shape of cells for both RRH and the C-RAN itself in D-RAN and C-RAN. The area of hexagonal based RRH then can be calculated as $A = \frac{r_c^2 \times 3 \times \sqrt{3}}{2} = 24\sqrt{3}\text{ km}^2$, while C-RAN (which its radius equals to 24 km according to previous

analysis) area is $A_c = 36 A$. This leads to each C-RAN to be composed of 36 RRHs. On the other side, the area of hexagonal based C-RAN is $A_c = \frac{R_c^2 \times 3 \times \sqrt{3}}{2}$.

TABLE I
DELAY COMPONENT VALUES

Delay component	Unit of delay	Table.1 entry	Required values
A. RT RF processing time	RRH	1,13	$\sim 40 \mu\text{sec}$
B. RT CPRI processing time	RRH,BBU	2,5,9,12	$\sim 10 \mu\text{sec}$
C. RT BBU processing time	BBU	6,7,8	$\sim 2700 \mu\text{sec}$
D. Fronthaul equipment's (if exist) delay	Fronthaul	4,10	$\sim 40 \mu\text{sec}$

III. PROPOSED C-RAN ARCHITECTURE

A. Related Work

The authors in [3] [4] have adopted a partial SDN approach based on decoupling the control and data planes at the serving gateway (SGW) which becomes an advanced OF switch that enables to encapsulate/de-encapsulate GPRS Tunnelling Protocol (GTP) packets while SGW-C has been transferred to the entity where the OF controller and MME reside. This approach is innovative and has shown an improvement in the total reduction of signalling load. However the new design has been considered as a partial solution as the PDN gateway (PGW) as the functional entity is following the existing 3GPP architecture. The authors in [5] have proposed a complementary vision which is fully realized in OpenFlow where the PGW-C has been decoupled and virtualized as an application running on top of the OF controller. This methodology is based on a complete separation between the control and data planes. For convenience, the proposed architectures in [5] are called OF DRAN and partial OF DRAN in [3][4]. Five main signalling procedures are evaluated in previous studies, Handover and tracking Area Update. The analytical results of adopting OF DRAN against the legacy topology & partial OF DRAN show an overall improvement in decreasing the number of messages being exchanged between all entities in an hour. It has been assumed that each UE supports multiple applications (K types) with certain arrival rate of λ_k , in addition, circular shaped cells of number C have been chosen and an area of consideration A, N_{ue} is total number of UEs and ρ_{ue} is the user density given by $\frac{N_{ue}}{A}$.

B. Proposed Archetype 1

The proposed architectures will analyse the signalling load based on the signalling analysis model proposed in [8]. The first architecture applies a major amendment on the architecture and utilizes the methodology of the Distributed MME signalling

load. For ease of use we will call it C-RAN.D-MME. This architecture is based on integration of the MME functionality within the BBU in the data centre. In this aspect each C-RAN is seen as single cell with its own BBU and MME pooling. One C-RAN is composed of a variable number of RRHs. The SGW-C, PGW-C and OF controller are combined and Packaged in one Entity called Centric Controller CC. Three main C-RANs are proposed as illustrated in Fig.1. The study will address all signalling messages being exchanged in all entities.

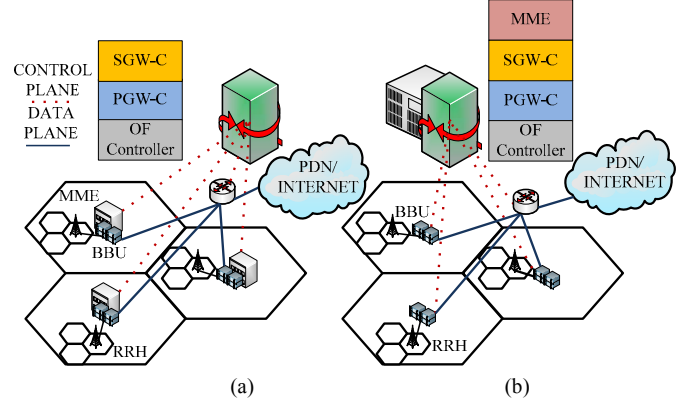


Figure. 1: Proposed CRAN with (a) Distributed MME & (b) Centralised MME

C-RAN.D-MME Initial attachment procedure is proposed to register the UE information and apply authorization and authentication plus creating sessions between the MME and SGW-C. The call flow is aligned with [9]. The total number of messages is equal to 8, where the probability that the UE initiates an attachment procedure in the network is assumed as $P_{initial} = 0.2$, thus the total signalling load $S_{initial}$ is

$$S_{initial} = P_{initial} \cdot A \cdot \rho_{ue} \cdot C \cdot 8. \quad (1)$$

The second procedure to address is the UE-generated services. This event takes place when the UE which is in IDLE state triggers the need to set up a connection with the PDN/Internet. In the standard LTE architecture, a few messages are exchanged between the eNB and the MME including the authentication check, initial context setup request and initial context setup response. However C-RAN.D-MME eliminates the need for such messages, and this will save the related RF and power resources. The BBU will be assessed as the eNB in former studies (OF enabled entity) thus whatever applies for the OF switch can be mapped in the same manner on it. That's why when the UE sends its first packet to the BBU, the BBU looks up in its flow table to find a matching rule. In case the entry is not found, an OF (packet-in) message is sent to the CC in order to investigate the message, acquire source and destination IP address and interact with SGW-C and PDW-C in an internal process to obtain the GW-U required information. Consecutively, the CC generates flow rules for subsequent packets. Each service k has an arrival rate λ_k (sessions/hour/UE) and average session duration of μ_k^{-1} . we denote P_k as the probability that session k is generated by the UE. The total number of messages in this case is 6, therefore, the total signalling for such an event is given by:

$$S_{uetrig} = \lambda_k \cdot P_k \cdot \rho_{ue} \cdot C \cdot A \cdot 6. \quad (2)$$

The third procedure is network triggered service, where in this analysis both cases of the UE being either in IDLE or CONNECTED state are taken into account. C-RAN.D-MME new architecture proposes some changes in the entities functionalities. Unlike the paging procedure in the standards [9], where paging the IDLE UEs is one of the MME functions, in our proposed architecture we are assuming that related connection management is one of CC's functions instead of MME. This implies when an incoming session triggers the CC, it will page all MMEs in its domain. This leads to unicast paging for a limited number of times (the same as number of C-RAN in its domain) instead of paging all eNBs. When the UE is in CONNECTED state, the number of exchanged messages is reduced by one, which is the paging. The total signalling load can be given by $S_{NWtriggered} =$

$$((8 + C_{ran}) \cdot R_p \cdot P_i + 7 \cdot (1 - P_i)) \cdot \lambda_k \cdot (1 - P) \cdot \rho_{ue} \cdot C \cdot A. \quad (3)$$

Where R_p is the average number of paging per transmission, C_{ran} is the number of C-RAN cells in the considered area. P_i is the probability of UE being in IDLE state which is computed from the process of (X_n, Y_n) as : $P_i = \prod_{n=1}^N \frac{\mu_n}{(\lambda_n + \mu_n)}$ according to the alternating renewal process [8], the UE is in CONNECTED state when it has at least one active session thus its related probability is: $P_{co} = 1 - P_i$.

The fourth procedure to analyse is the handover (HO). For simplicity, the HO will be divided to outer handover which occurs between CRANs and inner one that occur within CRAN itself considering mobility within the boundary of the CRAN. From Fig.2 it can be observed that 4 main messages are exchanged (Radio configuration and Radio configuration complete are considered as one message), hence the inner HO signalling load is computed by:

$$S_{l1hx2cran} = R_c \cdot (1 - P_i) \cdot C_{inran} \cdot C_{ran} \cdot 4. \quad (4)$$

where C_{inran} is the RRs number in single C-RAN and C_{ran} is the number of C-RAN cells. Their multiplication $C_{inran} \cdot C_{ran}$ results in total number of cells in the considered area. The mobile crossing rate of the enclosed area is $R_c = \frac{\rho_{ue} \cdot L \cdot V}{\pi}$ based on a fluid flow model [10]. L is the cell perimeter length, V is the UE average velocity. The outer HO calculation is assumed to be based only on the S1 interface between C-RANs; the crossing rate out in the outer HO is based on L_{cran} (the perimeter of a single C-RAN).

Hence the outer HO calculation is given by:

$$S_{l1hos13} = R_{cran} \cdot (1 - P_i) \cdot C_{ran} \cdot 9. \quad (5)$$

Since each C-RAN has its own MME, the HO that occurs between C-RANs can be considered as inter-MME HO. However we have only one SGW-C in our architecture, consequently there is no SGW-C relocation. Details of signalling procedures are not presented in this work as the purpose is showing the advantages of reshaping the network architecture rather than discussing signalling messages which are presented in detail in [9].

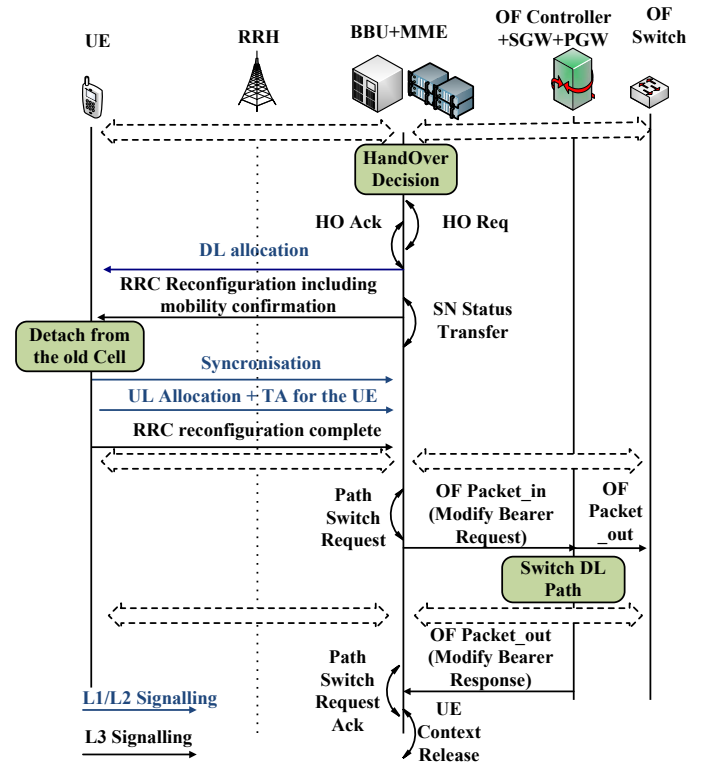


Figure. 2: Scenario 1 X2 HO call flow

The final important procedure is Tracking Area Update (TAU). This event is initiated when the UE moves and detects a new tracking area that is not in the tracking areas list allocated by the MME at the time of UE attachment or when the TAU timer expires. However we will ignore the second condition in this study. This procedure has different call flows in legacy LTE/EPC architectures depending on MME relocation or not and arises irrespective of whether the UE is in IDLE or CONNECTED state. It has one constant call flow in this proposed scheme, C-RAN.D-MME assumes that each C-RAN is a tracking area by itself served by its MME, no TAU occurs unless the UE crosses between C-RANs. The UE sends TAU to the RRH which passes it to the BBU. The MME is integrated and located at the same data centre as the BBU thus no signalling is needed between them. The target MME will then update the location of the UE to the home subscriber server (HSS) for future incoming sessions. A couple of messages are required to be exchanged (source and target MME) about cancelling the location information at the source MME and inserting subscription data at the target one. When the UE is in CONNECTED state, the target MME has to communicate with SGW-C for U-plane programming. The rate of crossing the tracking area is estimated by crossing out of a cell multiplied by $\frac{1}{\sqrt{C_{inran}}}$ where $\sqrt{C_{inran}}$ is the size of tracking area in this scheme. Concluding above the total signalling for TAU is:

$$S_{ltaudist} = \frac{1}{\sqrt{C_{inran}}} \cdot R_c \cdot C \cdot 10. \quad (6)$$

C. Proposed Archetype 2

The second proposed architecture (C-RAN.C-MME) resembles the first scheme in terms of C-RAN topology as demonstrated in Fig 1(b), nevertheless it considers a central MME rather than distributed. The functions of the central MME are virtualized as an application like SGW-C, PGW-C in C-RAN.D-MME where all of them run on top of the OF controller and communicate with its API. The rest of the network entities are kept the same. This section will only highlight the signalling load calculations as the main functions are illustrated in the previous part. The initial attachment signalling load in this scenario is

$$S_{initial} = P_{initial} \cdot A \cdot \rho_{ue} \cdot C \cdot 8. \quad (7)$$

The UE-triggered total signalling call flow is shown in Fig.3 given by:

$$S_{uetrig} = \lambda_k \cdot P \cdot \rho_{ue} \cdot C \cdot A \cdot 8. \quad (8)$$

The C-RAN.C-MME has only one central MME and the size of the tracking area is not constant but variable due to its configuration. When DL traffic needs to be delivered to the UE in the IDLE state (only its tracking area known to MME) the CC performs paging and sends unicast messages to all eNBs (RRHs) in its tracking area. The total signalling load for this procedure is given by:

$$S_{NWtriggered} = [((10 + C_{tau}) \cdot R_p \cdot P_i + 9 \cdot (1 - P_i)) \cdot \lambda_k \cdot (1 - P) \cdot \rho_{ue} \cdot C \cdot A]. \quad (9)$$

The total signalling for the X2 based inner HO is given by:

$$S_{l1hx2cran} = R_c \cdot (1 - P_i) \cdot C_{inran} \cdot C_{ran} \cdot 4. \quad (10)$$

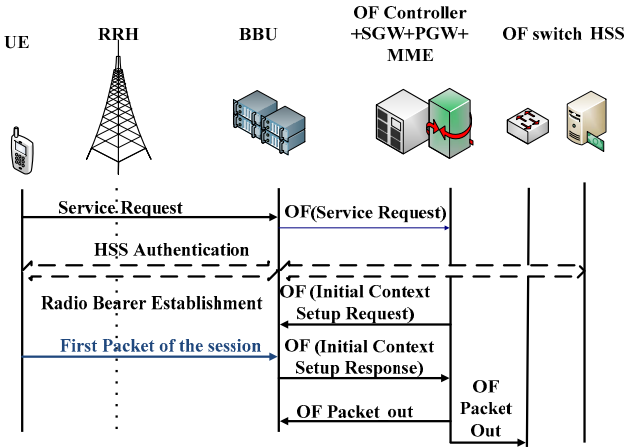


Figure 3: Scenario-2 UE-triggered service call flow

while total signalling due to outer HO is:

$$S_{l1hos13} = R_{cran} \cdot (1 - P_i) \cdot C_{ran} \cdot 9. \quad (11)$$

As there is only one virtualized MME in the CC, the tracking area update procedure is intra-MME only. In this architecture, the MME simply records the UE's new location and accepts the TAU, therefore there is no signalling to the HSS. The total signalling for TAU procedure is given:

$$S_{ltaudist} = \frac{1}{\sqrt{C_{tau}}} \cdot R_c \cdot C \cdot 6. \quad (12)$$

Tracking area update signalling is different to (C-RAN.D-MME) as it depends on the variable parameter C_{tau} .

IV. NUMERICAL AND ANALYTICAL RESULTS

In this section, we present the numerical results on signalling load for proposed (C-RAN.C-MME) & (C-RAN.D-MME) architectures, legacy network and OF DRAN architectures proposed in [5]. The calculations in this paper will only acknowledge one paging case which is unicast. ([5] [8] have investigated the optimality between multicast and unicast, thus it is sufficient in this study to use just one of them). The first case is based on increasing the area of the region; the area range to analyse spans between [800...3000] km^2 land excluding water. If the number of RRHs and C-RANs is constant, then increasing the area causes a proportional increase of RRH and the C-RAN cell's radius. It is also assumed that we have the four topologies forms that are presented in Table II where the total number of RRHs is 108 regardless what the form type is. Each form has different number of C-RANs and RRHs within to find the optimum one. In addition, all prior topologies comply with the timing constraint indicated in section II. The total signalling load is calculated by simple summation of (1, 2, 3, 4, 5, 6) equations and (7, 8, 9, 10, 11, 12) for C-RAN.D-MME, and C-RAN.C-MME respectively. The region to consider has 1 million users.

TABLE II
C-RAN FORMS

CRAN FORM	No. CRAN	No. RRH in one CRAN
CRAN 1	3 C-RANs	36 RRHs
CRAN 2	4 C-RANs	27 RRHs
CRAN 3	6 C-RANs	18 RRHs
CRAN 4	9 C-RANs	12 RRHs

The users are uniformly distributed. Only one application is taken into account with average arrival rate of $\lambda_k = 0.05$ and average session duration of 0.1. $P_k = 0.5$, $R_p = 1.1$. Uniform hexagonal cells are assumed with an overlapping factor of $\gamma = 1.2$, $C_{tau} = 6$, $V = 20 km/h$. Fig.4 highlights the total signalling load differences between all schemes considered earlier of C-RAN.D-MME (scenario 1), C-RAN.C-MME (scenario 2), OF DRAN and legacy. The evaluation is calculated based on the X2 interface for inner HO and S1 for outer HO. It's clear to observe that C-RAN.D-MME of CRAN1 topology experience the least amount of the load followed by the second form CRAN2. Nevertheless scenario 2 CRAN1 shows better performance compared to scenario 1 CRAN 3 & 4, OF DRAN and legacy network schemes. The signalling load in OF DRAN case is better than scenario 1 CRAN4. CRAN1 has the highest number of RRHs in a single CRAN, which means the lowest related TAU and outer HO signalling load. Furthermore increasing the area while keeping the number of users the same will definitely decrease signalling load which is a function of ρ_{ue} that is inverse proportion to the area A. Based on A, C, γ the cell radius is given by $r = \gamma \cdot \sqrt{\frac{2A}{3C\sqrt{3}}}$. Another metric to investigate is the TA size. Increasing the TA size will

have an impact on scenario 2 and OF-DRAN due to the decrease in TAU rate as Fig.5 illustrates. Nevertheless the C_{tau} increment alters the UE-terminated procedure as the latter is a linear function of it. The total signalling load keeps decreasing when increasing the TA size but on a very slow rate at high TA size. Meanwhile, scenario 1 doesn't require any alteration as the TAU size in this scenario is constant and equals to the number of RRHs in a single CRAN. Fig.5 further indicates that scenario 1 performs the optimum load compared to both scenario 2, OF DRAN and the legacy architecture by saving up to 17%, 36% and 62.8% respectively of signalling load for TA size of 6. Nonetheless, when the TA size raises beyond 10, scenario 2 becomes the most efficient architecture. It's observed that OF DRAN performs better than scenario1 when TA size is above 18. Moreover the higher the TA size, the less the load for our legacy architecture is. Both scenario 1 and the legacy scheme are observed to converge at very high TA sizes.

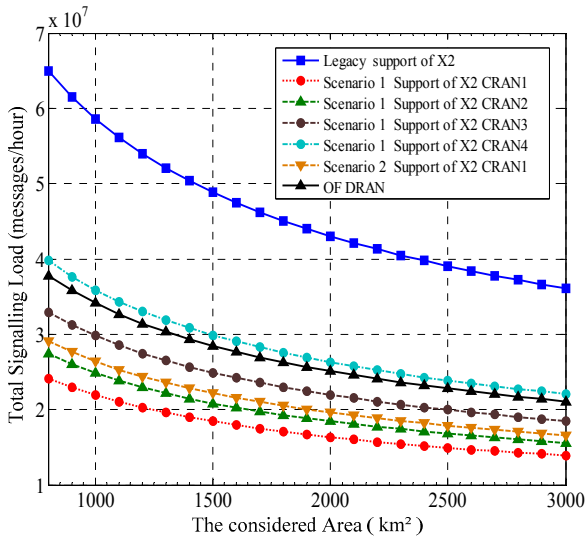


Figure. 4: Area based Comparison between Scenario 1 & 2

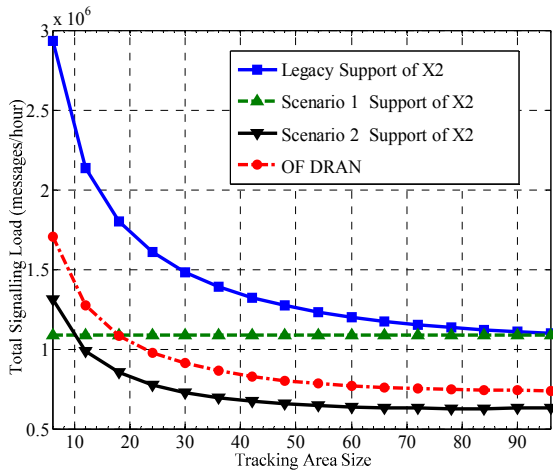


Figure.5: TAU size based Comparison between SC 1 & SC 2

V. CONCLUSION

In this paper we presented new C-RAN architectures by utilizing both SDN LTE architecture and C-RAN schemes. Two new design schemes have been proposed which are: C-RAN.C-MME and C-RAN.D-MME architectures in order to reduce overall control signalling load. The paper considered multiple network parameters such as cell area and tracking area update size. The proposed archetypes are shown to offer improvement in the overall signalling load as compared to the existing literature and previous suggested topologies. The impact of tracking area size on system performance has been investigated. We observed from our analysis that for small TA sizes, the C-RAN.D-MME gives the best performance while C-RAN.C-MME is better for higher TA sizes. Our results show that even different C-RAN topologies within the same architecture have different signalling loads.

REFERENCES

- [1] Tellabs, "Tellabs "end of profit" study executive summary," Tech. Rep., January 2011.
- [2] Open Networking Foundation (ONF), Software-defined networking: The new norm for networks. ONF White paper (2012). [Online] <https://www.opennetworking.org/images/stories/downloads/sdn-resources/white-papers/wpsdn-newnorm.pdf>.
- [3] Ben Hadj Said, S.; Sama, M.R.; Guilloard, K.; Suci, L.; Simon, G.; Lagrange, X.; Bonnin, J.-M., "New control plane in 3GPP LTE/EPC architecture for on-demand connectivity service," in *Cloud Networking (CloudNet), 2013 IEEE 2nd International Conference on*, vol., no., pp.205-209, 11-13 Nov. 2013.
- [4] Sama, M.R.; Ben Hadj Said, S.; Guilloard, K.; Suci, L., "Enabling network programmability in LTE/EPC architecture using OpenFlow," in *Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt), 2014 12th International Symposium on*, vol., no., pp.389-396, 12-16 May 2014.
- [5] VG Nguyen, Y Kim "Proposal and evaluation of SDN-based, Mobile packet core networks", J Wireless Com Network, 2015.
- [6] Liang Liu; Feng Yang; Wang, R.; Zhenning Shi; Stidwell, A.; Daqing Gu, "Analysis of handover performance improvement in cloud-RAN architecture," in *Communications and Networking in China (CHINACOM), 2012 7th International ICST Conference on*, vol., no., pp.850-855, 8-10 Aug. 2012.
- [7] Harrison J. Son and S.M.Shin. (April 2014). Fronthaul Size: Calculation of maximum distance between RRH and BBU.[online] <http://www.netmanias.com/en/?m=view&id=blog&no=6276>.
- [8] Widjaja, I.; Bosch, P.; La Roche, H., "Comparison of MME Signalling Loads for Long-Term-Evolution Architectures," in *Vehicular Technology Conference Fall (VTC 2009-Fall), 2009 IEEE 70th*, vol., no., pp.1-5, 20-23 Sept. 2009.
- [9] 3GPP, "General Packet Radio Service (GPRS) enhancements for Evolved Universal Terrestrial Radio Access Network (E-UTRAN) access," 3rd Generation Partnership Project (3GPP), TS 23.401, Jun. 2011.
- [10] D. Lam, D.C. Cox and J. Widom, "Teletraffic Modeling for Personal Communications Services," in *Communications Magazine, IEEE*, vol.35, no.2, pp.79-87, Feb 1997.